

# 变频调速水泵智能故障诊断方法

舒服华

(武汉理工大学机电工程学院,湖北 武汉 430070)

**摘要:**提出了一种粗糙集理论与神经网络集成的调速水泵故障诊断方法。针对调速水泵运行非恒速度的特点,首先用基于数字重采样的阶比分析方法对水泵振动信号进行预处理,然后利用粗糙集理论进行故障诊断决策系统约简,获得最优决策系统,最后在最优决策系统的基础上,设计 RBF 神经网络对调速泵进行故障诊断。试验结果显示,该方法能有效提高变频调速水泵故障诊断的精度和效率。

**关键词:**粗糙集;神经网络;阶比分析;调速水泵;故障诊断

**中图分类号:** TP118      **文献标识码:** A

## 0 引言

水泵是农业生产中常用的设备,在抗旱排涝、农田水利灌溉等方面发挥了重要作用。而变频调速水泵以其显著的节能效果在农业生产中的应用也越来越广泛。变频调速水泵的故障诊断与非变频调速水泵的故障诊断方法有所不同。非变频调速水泵的转速变化不大,近似恒定,因此可以认为其振动是一个平稳过程,通过进行频域分析就可以判断故障原因。但变频调速水泵的速度经常发生变化,振动是一个非平稳过程,频域分析很难识别故障原因<sup>[1]</sup>。因此寻求先进的技术进行变速泵故障诊断具有重要意义。人工神经网络(ANN)具有大规模的并行处理和分布式的信息存储能力、良好的自适应性、自组织性及很强的学习、联想、容错和抗干扰能力,能以任何精度逼近复杂非线性系统,在故障诊断领域得到了广泛的应用。粗糙集理论(Rough Sets, RS)是一种处理模糊性与不确定性的软计算基础理论,它不需要任何先验知识,仅利用数据本身提供的信息就能够表达和处理不完备信息,能在保留关键信息的前提下对数据的属性进行约简并求得知识的最小表达,从中发现隐含知识,揭示潜在的规律<sup>[2,3]</sup>。本文提出了一种 RS 理论和 ANN 融合的变速泵故障诊断方法。利用阶比分析方法的数字重采样对水泵振动信号进行预处理,以各频段信号的能量来构建故障特征向量,应用粗糙集理论对故障诊断决策系统进行约简,消除冗余信息,构建 RBF 神经网络,优化网络结构,从而提高网络的训练速度和故障诊断精度。

## 1 阶比分析数字重采样

传统的傅立叶变化是以等时间间隔采样的,不能跟踪变频调速水泵的转速变化,若用傅立叶变化的方法进行频域分析,必然产生较大的误差,也会造成其频谱分散、模糊和幅值度量不准等问题。阶比分析方法的数字重采样是解决这一问题的有效途径。阶比分析的实质是将时域的非平稳信号通过恒定角增量采样转变为角域的稳定信号,使其更好地反映与转速相关的振动。而采用阶比谱分析的方法,用等转角采样代替等时间采样,将等转角采样所得的离散转角函数经过离散傅立叶变化后就得到阶比谱函数。阶比代表每转的循环次数(也可理解为对应转速的倍数。例如某信号的振动频率为对应转轴旋转频率的 4 倍,则其阶比为 4),它可很好地表示和转速有关的振动。由于一般的 DFT 都是采用 2 倍的整数次幂取点,因而只要每转的采样数也取 2 的整数次幂,就可以保证周期采样。这样即使转速在一定的范围内波动,也可减小阶比分量的误差。数字重采样是将异步采样得到的振动和转速信号通过数据信号处理算法用软件的形式合成同步采样振动数据。基于数字重采样的阶比分析的步骤如下:①对原始振动信号进行等时间间隔采样和对健相信号进行等角度间隔采样,得到异步采样信号。②通过采集的健相脉冲序列,进行转速估计,并作为振动相角的测量基准。③由转角估计计算等转角采样发生的时间序列。④在计算出的等角度采样发生时刻附近时间区间内,对异步采样振动信号进行插值计算,最后重采样生成同步采样信号。

## 2 粗糙集知识约简

在粗糙集分类中,一些条件属性是必要的,而另一些可能是多余的,去除这些多余的属性不会影响原来的分类效果。对故障数据预处理,就是利用粗糙集的属性约简去除冗余的属性条件,简化神经网络的输入。具体步骤为:①对输入信息进行离散化。②运用粗糙集理论对原始数据进行约简。③从多种简化中选择一个特征子集。

### 2.1 数据离散化

RS只能处理离散属性值,而原始故障诊断系统的属性值是连续的,因此必须对决策系统中的数据进行离散化处理,本文采用自组织映射神经网络方法。它是一种无监督自组织竞争学习型前馈神经网络,能通过自组织方式利用大量的训练样本数据来调整网络权值,分类结果较为客观反映数据的实际分布情况。具体步骤如下<sup>[1]</sup>:①给定SOM网络初始权值 $m=2$ 。②给定条件属性初始维数 $d=1$ 。③将第 $d$ 维条件属性按大小排序。④用SOM网络对排序后的第 $d$ 维条件属性值进行分类。⑤把相邻两类边界属性值的均值作为二类的分界值。⑥用条件属性量化参数对该维条件属性进行量化。⑦赋值 $d=d+1$ ,返回步骤③直到最后一维条件属性。⑧检查数据表是否相容,若相容,停止,否则令 $m=m+1$ ,返回步骤②。

### 2.2 知识属性约简

属性约简基本思想是采用某种衡量标准确定不同属性的重要程度,构造最小子集。基于RS理论的知识约简方法分两个步骤,一是从决策表删除一些列;二是删除冗余的行。约简的方法比较多,但算法的代价比较高,本文采用遗传算法对属性进行约简,它具有比较好的通用性、鲁棒性、全局搜索性<sup>[3]</sup>。

(1)参数编码。编码形式采用二进制,基因的每一位代表区分矩阵的一项,即两个对象的区分属性集,某位0表示该属性不存在,1表示该属性存在。每个基因的长度等于条件属性的个数 $N$ 。

(2)适应度函数。决策约简的目标是用尽量少的属性来区分尽可能多的个体,为了达到这一目的,适应度函数设计为<sup>[1]</sup>:

$$F(x) = \frac{N-L(x)}{N} \times P - \frac{C(x)}{M} \times Q \quad (1)$$

式中: $N$ 为决策表中条件属性的个数; $L(x)$ 为候选约简 $x$ 所涉及的属性数量; $C(x)$ 为采用的约简方法能区分的个体数目; $M$ 为决策表中所包含的个体总数; $P$ 、 $Q$ 为比例系数, $P+Q=1$ 。

式中第一项的含义是希望用尽量少的属性实现约简,第二项的含义是希望约简方法能够区分尽可能多的个体。

### 2.3 最小属性子集评价

约简是不含多余属性并保证分类正确的最小条件属性集,而核是影响分类的重要属性,一般情况下,属性约简不是唯一的,而核是唯一的。虽然最小约简中包含的属性数目相同,但是,由不同的约简得到的规则数目及其适用范围有所差别,实际应用中,一般希望决策系统中的规则数目越少越好,而每条规则的适用范围越大越好,即条件属性的每一聚类中能够包含的对象较多,并且产生较少的规则,该指标可以通过聚类比来描述<sup>[3]</sup>:

$$R_c = \frac{N_0 - N_r}{N_0 - 1} \quad (2)$$

式中: $N_0$ 为原始决策系统中对象的数目; $N_r$ 为约简后决策系统中规则的数目。

## 3 径向基函数(RBF)神经网络

BP神经网络存在收敛速度慢和容易陷入局部最小点等缺点,因此本文选用径向基函数(RBF)神经网络作为故障诊断模型。RBF网络的输入节点只是传递信号到隐层,隐含层由类似高斯核函数那样的辐射状作用函数构成,而输出节点通常是简单的线性函数。由于输入到输出是非线性的,而网络输出对可调参数的映射又是线性的,网络的权值可递推计算,从而大大加快了学习速度并避免了局部极值问题。其数学模型为:

$$y = \sum_{i=1}^{n_c} \omega_i g(\|x - c_i\|, \sigma_i) \quad (3)$$

式中: $x$ 为神经网络输入; $\omega_i$ 为输出层权值; $g(\cdot)$ 为径向基函数; $c_i$ 为径向基函数的中心; $\sigma_i$ 为径向基函数的宽度; $n_c$ 为隐含层神经元数; $\|\cdot\|$ 为输入 $x$ 和 $c_i$ 之间的距离。

对于训练样本 $D_M = \{x_i, y_i\}, i=1, \dots, M$ ,寻找参数 $\Theta = \{c_i, \sigma_i, \omega_i, n_c\}$ ,使误差 $E$ 最小。

$$E = \frac{1}{2} \sum_{i=1}^{n_c} |y_i - f_n(x, \Theta)|^2 \quad (4)$$

式中:

$$f_n(x, \Theta) = \sum_{i=1}^{n_c} \omega_i g(\|x - c_i\|, \sigma_i) \quad (5)$$

RBF网络的学习过程分为两个阶段:①根据所有训练样本决定隐含层神经元 $n_c$ 及各节点的径向基函数的中心值 $c_i$ 和径向基函数的宽度 $\sigma_i$ ,利用聚类算法对学习样本聚类,得到RBF径向基函数的中心参数。并用自动终止聚类判断确定隐含层神经元数目。②在确定好隐含层的参数后,根据训练样本,利用递推最小二乘法(RLS)原则,求出输出层的权值 $\omega_i$ 。

## 4 诊断实现分析

对某型号的变频调速水泵进行诊断试验。运用加速度传感器测量水泵轴的振动信号。根据上述方法得到水泵的阶比波谱,取振动阶比波谱信号中 $(0 \sim 1/4)f$ ,  $(1/4 \sim 3/4)f$ ,  $1f$ ,  $2f$ ,  $3f$ 的频率段能量为故障特征参数(分别用 $x_1, x_2, x_3, x_4, x_5$ 表示, $f$ 为水泵轴的旋转频率),诊断水泵转子不平衡,轴承油膜振动,转子不对中故障(分别用 $y_1, y_2, y_3$ 表示),获得故障数据样本25组,取16组为训练样本见表1(数据经过加权优化处理),9组为测试样本。用SOM网络方法对连续属性进行离散化,分类数为3,离散结果如表2。用遗传算法对表2决策系统进行约简,得最小约简子集: $\{C_2, C_3, C_4\}$ ,  $\{C_1, C_3, C_5\}$ ,  $\{C_1, C_3, C_4\}$ 。经计算,约简 $\{C_2, C_3, C_4\}$ 的聚类比 $R_c=0.5625$ 为最大,选其取为最优决策系统如表3。建立RBF神经网络,输入层神经元数为3,分别对应约简后得3个条件属性,输出层神经元数为3,分别对应水泵的3种故障类型,隐含层神经元由训练样本和算法确定,训练完成后,用测试样本进行测试,结果如表4,诊断结果与实际情况完全一致。

表 1 训练样本数据

样本	状态	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	$y_1$	0.022 5	0.101 4	0.720 4	0.141 6	0.015 0
2	$y_2$	0.005 6	0.603 2	0.313 2	0.115 6	0.021 1
3	$y_3$	0.025 1	0.122 9	0.513 8	0.341 5	0.001 0
4	$y_2$	0.014 0	0.470 1	0.352 2	0.147 4	0.016 3
5	$y_2$	0.009 0	0.631 5	0.221 5	0.102 1	0.016 9
6	$y_1$	0.016 7	0.130 0	0.624 2	0.218 4	0.012 7
7	$y_1$	0.015 6	0.100 2	0.692 3	0.181 6	0.010 3
8	$y_1$	0.009 5	0.109 6	0.660 1	0.211 4	0.009 4
9	$y_3$	0.016 1	0.114 4	0.520 1	0.301 9	0.026 5
10	$y_3$	0.014 9	0.090 4	0.601 7	0.290 5	0.002 5
11	$y_3$	0.004 1	0.154 3	0.513 2	0.313 2	0.015 2
12	$y_2$	0.008 2	0.550 1	0.259 1	0.162 1	0.020 5
13	$y_1$	0.011 6	0.069 4	0.790 6	0.181 1	0.018 6
14	$y_1$	0.005 1	0.102 0	0.687 1	0.190 9	0.014 9
15	$y_2$	0.002 5	0.406 0	0.470 9	0.128 6	0.002 8
16	$y_3$	0.018 6	0.131 2	0.521 1	0.290 3	0.021 0

表 2 数据离散化结果

$U$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$D$
1	3	1	3	1	2	1
2	1	3	2	1	3	2
3	3	1	2	3	1	3
4	2	3	2	1	3	2
5	1	3	1	1	3	2
6	3	1	3	2	2	1
7	3	1	3	2	2	1
8	1	1	3	2	1	1
9	3	1	2	3	3	3
10	2	1	3	3	1	3
11	1	2	2	3	2	3
12	1	3	1	1	3	2
13	2	1	3	2	3	1
14	1	1	3	2	2	1
15	1	3	2	1	1	2
16	3	1	3	2	3	3

表 3 最优决策系统

$U$	$C_3$	$C_4$	$C_5$	$D$
1	1	3	1	1
2	3	2	1	2
3	1	2	3	3
5	3	1	1	2
6	1	3	2	1
10	1	3	3	3
11	2	2	3	3

表 4 测试样本及测试结果

样本	状态	$C_3$	$C_4$	$C_5$	$y_1$	$y_2$	$y_3$	诊断结果
1	1	1	3	2	0.995 9	0.000 1	0	不平衡
2	1	1	3	1	0.979 8	0.000 5	0.002 1	不平衡
3	1	1	2	2	0.989 8	0.001 5	0	不平衡
4	2	2	2	1	0.000 2	0.969 7	0.005 4	油膜振动
5	2	2	2	1	0.000 2	0.969 7	0.005 4	油膜振动
6	2	3	1	1	0.015 1	0.997 6	0.025 5	油膜振动
7	3	1	3	3	0.000 7	0	0.996 3	不对中
8	3	2	2	3	0.011 7	0.000 6	0.945 8	不对中
9	3	1	3	3	0.000 7	0	0.996 3	不对中

为了检验本文粗糙集与神经网络融合的性能,设计一个 BP 神经网络,直接利用原始数据进行故障诊断,网络结构 5—8—3。两套神经网络模型的基本参数比较如表 5(ANN1 和 ANN2 分别代表以约简前、后数据作为训练样本的神经网络模型)。从表 4 可见,ANN2 与 ANN1 相比,输入单元,样本数和训练步数都有较大减少,而 ANN2 的拟合残差,训练时间却比 ANN1 小许多。用测试样本对训练好的 ANN1 进行测试,虽然也全部识别正确,但精度较 ANN2 低,并且处理时间较长。这表明基于粗糙集约简处理的 ANN 简化了网络的结构,显著提高了网络学习效率和诊断精度。

表 5 神经网络模型比较

网络模型	输入单元	隐层单元	输出单元	样本数量	学习部长	训练部数	训练时间/s	拟合残差
ANN1	5	10	3	16	35	985	1.56	0.01
ANN2	3	4	3	7	20	472	0.02	0.000 1

## 5 结 语

(1)采用阶比分析方法的数字重采样对水泵振动信号进行预处理,可以有效地提高拾起的振动信号的精度,准确地跟踪水泵转速的变化。

(2)利用粗糙集理论对故障诊断决策系统进行约简,降低了输入向量的维数和训练样本的数量,缩短网络的训练时间,提高泛化能力。

(3)RBF 神经网络可使网络结构优化,调整参数少,收敛速度快,网络的适用性强,具有较强的分类能力。

(4)粗糙集与神经网络的故障诊断方法,有效解决了变频调速水泵故障中,故障信号重叠干扰,故障与征兆的关系非线性和不明确等问题,提高了水泵故障诊断的可靠性和效率。

### 参考文献:

- [1] 傅其凤,崔彦平,葛杏卫. 变频调速水泵状态实时监测与故障智能诊断系统的研究[J]. 煤矿机械,2005,28(6):143—145.
- [2] 谢 川,倪世宏. 基于粗糙集理论的飞行数据模式特征提取[J]. 计算机工程,2005,31(6):169—171.
- [3] 李增芳,何 勇. 基于粗糙集理论与 BP 神经网络的发动机故障诊断模型[J]. 农业机械学报,2005,36(8):118—121.